

Anexo

Sergio Alejandro Esparza González: Se compararon distintas variables entre tres entidades federativas de México, para saber cuántas personas finalizan su educación posbásica después de los 18 años. Los estados elegidos fueron: Sonora, Estado de México y Veracruz.

Posteriormente, se seleccionaron las variables más importantes del dataset y se eliminaron los valores nulos de cada una. Se realizó una visualización de datos para las variables numéricas y categóricas, por ejemplo, gráficos de barras, histogramas de frecuencias, gráficas de pastel, gráficas de cajas, y así detectar los valores atípicos. Finalmente, se guardó el nuevo dataset en un nuevo archivo con los datos ya listos para análisis futuros.

En este proyecto acerca de la educación posbásica, se llevaron a cabo varios análisis que tuvieran impacto para su realización, por ejemplo, la selección de variables para su posterior limpieza y preparación para el análisis utilizando notebooks de Google Colaboratory con el lenguaje de Python. Después se ajustaron distintos modelos de aprendizaje automático utilizando distintos algoritmos como K-Nearest Neighbors, Decision Trees y Random Forests. Con esta experiencia, pudimos validar las hipótesis iniciales para los factores del éxito educativo usando modelación estadística y ciencia de datos. Se tomaron métricas de desempeño para evaluar el mejor algoritmo y se hicieron comparaciones acerca de los resultados para tomar las mejores decisiones posibles.